© 2012 г. **ALAR LEIBAK**, Ph.D., Associate Professor,
(Tallinn University of Technology, Department of Mathematics),
ANNA ŠELETSKI, Ph.D., Researcher,
OTU VAARMANN, Ph.D., Senior Researcher, Professor Emeritus,
(Institute of Cybernetics at Tallinn University of Technology)

ON A MULTI-LEVEL APPROACH TO THE GENERATION OF PARETO POINTS FOR COMPLEX SYSTEMS

For solving large and multi-objective optimization problems a multi-level approach to the generation of Pareto points is discussed. Frequently, multi-objective optimization problems are solved by the scalarization or weighting method where multiple criteria are replaced by a single function based on the weighted sum. According to the idea of hierarchical approach the overall complex problem is reduced through the introduction of coordination parameters into smaller and simpler subproblems which thereafter will be solved independently as the coordinated (harmonized) ones to generate the Pareto optimal points for the primary global problem. For finding proper values of coordination parameters some iterative methods based on the Gauss-Newton method are studied. Convergence properties and computational aspects of the methods for the evaluation of coordination parameters are discussed.

ON A MULTI-LEVEL APPROACH TO THE GENERATION OF PARETO POINTS FOR COMPLEX SYSTEMS / Alar Leibak (Tallinn University of Technology, Department of Mathematics Ehitajate tee 5, 19086 Tallinn, Estonia, E-mail: alar.leibak@ttu.ee), Anna Šeletski (Akadeemia tee 21, 12618 Tallinn, Estonia, E-mail: annatar@ioc.ee), Otu Vaarmann (Institute of Cybernetics at Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia, E-mail: vaarmann@ioc.ee)

1. Introduction

Every year optimization algorithms are called on to handle problems that are much more larger and complicated than that in the past. A number of problems in economics, business, engineering and scientific computation (e.g. production planning, process control, image restoration, logistics, neural networks, inverse problems etc.) lead frequently to a intricate mathematical problem. Many of them are not highdimensional but also of the multi-objective type. Public section decision-making (DM) typically involves complex optimization problems from very different backgrounds with several conflicting goals and there are compromises to be made between economic development and environmental quality in order to assess the ongoing sustainable management. Many multiple criteria decision problems can be posed as a multi-objective optimization problem.

Uncertainty, subjectivity, risk and conflicting views are inevitable circumstances in planning and decision making. With a wide range of causes and types of uncertainty, there are correspondingly many approaches to their treatment in decision analysis and optimization models. The principle of 'consensus' form the basis of the rules to assess the real-life problems and solutions obtained on this principle are known as Pareto optimal or efficient solutions. In case of Pareto optimality the trade-off solutions of interest are the non-dominant ones, i.e. solutions for which any attempt to improve their optimality with respect to one and more objectives results in a decrease in optimality with respect to one or more other objectives. Methods based on Multiple Criteria Optimization enable calculate or approximate the associated Pareto set and Multiple Criteria Analysis provides powerful and useful techniques for appraisal of complex decision problems yielding a good overall understanding of alternatives.

The multicriteria analysis is aiming at providing a formal approach helping decision makers to handle effectively complex decision situations such as climate change mitigation tools, energy planning etc. in which the level of conflicts between criteria is such that intuitive solution(s) can not be satisfactory. The energy-economy-environment modelling involve many sources of uncertainty therefore economic, socio-political, technical and environmental criteria should be considered in order that the decision maker can decide which projects are mature enough to be further promoted.

One of the widely used approaches to solve the multi-objective optimization problem is that to transform it into a problem with the single objective function. There are several ways of reducing a multicriterial problem into a monocriterial one. Many multiple criteria decision aiding methods use coefficients for representing trade-offs between conflicting objectives and for achieving agreements between conflicting parties. Attribution of numerical values to coefficients (weights) showing the relative importance of the different criteria reflects, as a rule, the subjective opinion of the decision maker, i.e. always implies some degree of arbitrariness and uncertainty. Therefore, sensitivity is an important complement to multicriteria analysis, since it investigates how the output changes when the judgements of the decision makers are varied. The multiple criteria problem is iterated by selecting iteratively different weight coefficients of the criteria. Thus the varying of these weights can be considered as a part of sensitivity analysis.

Frequently, multi-objective optimization problems are solved by the scalarization method known also as the weighting method. The aim is then to find the Pareto point for the numerical values of target functions $f^1(x), ..., f^n(x)$ with the argument x belonging to a certain Hilbert space. As the space of parameters is the same for all the functions, however, as a rule, the best points of target functions are different. Thus, in this case, certain compromise points must be found in the space of parameters. The idea of the weighting method is to associate each objective with a weighting factor and after that optimize the weighted sum of objectives, i.e. multiple criteria is replaced by an overall function $\sum_{i=1}^{n} \alpha_i f^i(x)$, where $\alpha_1, ..., \alpha_n$ are the non-negative coefficients with the constraint $\sum_{i=1}^{n} \alpha_i = 1$. The role of weighting coefficients may play the dual variables, the probabilities of appearance the criteria $f_i(x)$ etc. For some scalarization methods the Pareto optimality is guaranteed [1].

Many nonlinear optimization problems have a typical structure which can be exploited to facilitate and accelerate the computational process for solving them. Problems of high dimension, e.g. those, arising in electrical, mechanical, and aeronautical engineering can sometimes only be solved by the exploitation of their special structure [3]. The generation of either optimal or Pareto optimal points for large and complex systems can be easier if the problem is decomposable, i.e. it can be decomposed and solved as a set of smaller coordinated subproblems which can thereafter be distributed over a large number of processors and treated then independently enabling thus to organize parallel computation. Decomposition techniques are applicable provided the problem under consideration has a small number of linking constraints and linking variables. This approach is particularly successful for separable and partially separable problems [2] - [5]. Since decomposability is a kind of organized sparseness then the problem variables can be divided into groups so that most variables interact only with members of their own group. It should then be possible to solve the problem hierarchically: on the top level, setting values for a small number of variables common to the groups; and the lower level, solving independently within each group for those variables that interact only with others in the same group. Thus, in order to cope with large scale problems and to develop many optimum plans a multi-level (hierarchical) approach may be useful. One way to break the given overall problem into subproblems is the use of decomposition-coordination schemes [2] - [5] i.e. designating processors as master and slaves. The two most widely used decomposition-coordination schemes are feasible and non-feasible solutions ones. In the latter only the values reached at the end of the procedure are assured to be feasible. In the latter case, intermediate problems to be solved are also simpler and one can make use of the potential of the duality theory. Besides, the problem to be solved in the coordination phase is then an unconstrained one [2] - [5]. Here merely non-feasible solutions approach will be discussed.

2. Non-feasible solutions approach

In this section we consider the problem of generating efficient (Pareto optimal) solutions for a multi-objective optimization problem. If the hierarchical approach based on the computation of non-feasible solutions is implemented then the solution of the original problem is obtained by the maximization of the dual objective in the Lagrange-multiplier space. Note, that under certain conditions the set of efficient solutions for the original problem can be generated by the efficient solutions of the subsystems [4].

Next, we handle a multi-objective programming problem with convex, inequality-constrained subproblems and make distinction between linear and nonlinear constrains.

To demonstrate this approach of a problem solving based on the generation of non-feasible Pareto points we first present a formal description of a complex system taken from an excellent paper [4].

Consider a system N $(N \geqslant 2)$ interconnected subsystems S_i, $i = 1, ..., N$, with multiple objectives in each of them. Let $\mathbf{m}_i \in^{n_{m_i}}$, $i = 1, ..., N$, be the decision vector of S_i, $\mathbf{y}_i \in^{n_{y_i}}$ its output vector and $\mathbf{x}_i \in^{n_{x_i}}$ its input vector. Vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ generate the interconnection among the subsystems. By $\mathbf{k}_i^T = (\mathbf{x}_i^T, \mathbf{m}_i^T, \mathbf{y}_i^T)$ is denoted the vector of subsystem of dimension $n_{k_i} = n_{x_i} + n_{m_i} + n_{y_i}$ and by $\mathbf{k}^T = (\mathbf{k}_1^T, \mathbf{k}_2^T, \ldots, \mathbf{k}_N^T)$ the joint vector of variables of global problem of dimension $n_k = \sum_{i=1}^{N} n_{k_i}$. If to denote by $\mathbf{f}^i = (f_1^i, f_2^i, ..., f_{n_i}^i)^T$ the objective vector of subsystem i and by $\mathbf{f}^T = ((\mathbf{f}^1)^T, (\mathbf{f}^2)^T, ..., (\mathbf{f}^N)^T)^T$ the joint objective vector of all subsystems of dimension $n = \sum_{i=1}^{N} n_i$ then the multiple objective programming problem corresponding to the global system can be expressed as follows

$$\text{(1)} \qquad \text{minimize} \begin{bmatrix} \mathbf{f}^1(\mathbf{k}_1) \\ \mathbf{f}^2(\mathbf{k}_2) \\ . \\ . \\ . \\ \mathbf{f}^N(\mathbf{k}_N) \end{bmatrix}$$

subject to

$$\text{(2)} \qquad \mathbf{y}_i = \mathbf{A}_i(\mathbf{x}_i, \mathbf{m}_i) \quad i = 1, 2, ..., N$$

$$\text{(3)} \qquad \mathbf{g}_i(\mathbf{k}_i) \leqslant 0 \quad i = 1, 2, ..., N$$

$$\text{(4)} \qquad \mathbf{x}_i = \sum_{j=1}^{N} \mathbf{C}_{ij}\mathbf{y}_j \quad i = 1, 2, ..., N.$$

Equations (4) represent connections among the subsystems and indicate that the impact vector of each subsystem is a linear combination of the outputs of all the N subsystems. The connection matrix $\mathbf{C}_{ij}$ is usually constant and formed by zero-one elements, where the unity indicates a connection. Further on, it is assumed that function $\mathbf{f}^i$ and $\mathbf{g}_i$ are convex, functions $\mathbf{A}_i$ are linear and all functions are continuously differentiable.

In accordance with the idea of weighting method for solving multi-objective problems the vector function $\left((\mathbf{f}^1)^T, (\mathbf{f}^2)^T, \ldots, (\mathbf{f}^N)^T\right)^T$ in (1)-(4) is replaced by a sum of subproblems $\sum_{i=1}^{N} (\boldsymbol{\omega}^i)^T \mathbf{f}^i(\mathbf{k}_i)$, where $(\boldsymbol{\omega}^i)^T = (\omega_1^i, \omega_2^i, ..., \omega_{n_i}^i) \geqslant 0$ is the vector of weights associated to objective $\mathbf{f}^i$, $i = 1, 2, ..., N$. Under an additional constraint qualification condition ruling out certain pathological properties of the constraint set, there exists the dual function of weighted problem, with respect to the constraints (4), which is given by

$$\text{(5)} \qquad \Phi(\boldsymbol{\lambda}) = \min_{k \in \prod_{i=1}^{N} \mathcal{X}_i} \left\{ \sum_{i=1}^{N} (\boldsymbol{\omega}^i)^T \mathbf{f}^i(\mathbf{k}_i) + \sum_{i=1}^{N} \left[(\boldsymbol{\lambda}^i)^T \left(\mathbf{x}_i - \sum_{j=1}^{N} \mathbf{C}_{ij}\mathbf{y}_j \right) \right] \right\},$$

where $\mathcal{X}_i = \left\{ \mathbf{k}_i^T = (\mathbf{x}_i^T, \mathbf{m}_i^T, \mathbf{y}_i^T) \mid \mathbf{y}_i = \mathbf{A}_i(\mathbf{x}_i, \mathbf{m}_i),\ \mathbf{g}_i(\mathbf{k}_i) \leqslant 0 \right\}$, $i = 1, 2, ..., N$ and $\boldsymbol{\lambda} = ((\boldsymbol{\lambda}^1)^T, (\boldsymbol{\lambda}^2)^T, \ldots, (\boldsymbol{\lambda}^N)^T)^T$ and $\boldsymbol{\lambda}^i = (\lambda_1^i, \lambda_2^i, \ldots, \lambda_{n_{x_i}}^i)^T$ are the dual variables and they are associated to the coupling constraints (4).

As to constraint qualification condition the linear independence regularity condition may be used. It is known [6], if at least one objective function of the problem is strictly convex, then, Φ is differentiable at $\overline{\lambda} \in^{n_{\mathcal{X}}}$ with the gradient

$$\text{(6)} \qquad \nabla\Phi(\overline{\boldsymbol{\lambda}}) = \left(\left(\mathbf{x}_1(\overline{\boldsymbol{\lambda}}) - \sum_{j=1}^{N} \mathbf{C}_{1j}\mathbf{y}_j(\overline{\boldsymbol{\lambda}}) \right)^T, ..., \left(\mathbf{x}_N(\overline{\boldsymbol{\lambda}}) - \sum_{j=1}^{N} \mathbf{C}_{Nj}\mathbf{y}_j(\overline{\boldsymbol{\lambda}}) \right)^T \right)^T.$$

Multiplying and dividing by $\mu_i = \sum_{j=1}^{N} \omega_j^i$, $i = 1, 2, ..., N$, the second term in (5) can be written as

$$\text{(7)} \qquad \sum_{i=1}^{N} \left[(\boldsymbol{\lambda}^i)^T \left(\mathbf{x}_i - \sum_{j=1}^{N} \mathbf{C}_{ij}\mathbf{y}_j \right) \right] = \sum_{i=1}^{N} \left[\sum_{j=1}^{n_i} \omega_j^i \left(\frac{1}{\mu_i} (\boldsymbol{\lambda}^i)^T \mathbf{x}_i - \sum_{l=1}^{N} \frac{1}{\mu_i} (\boldsymbol{\lambda}^l)^T \mathbf{C}_{lj}\mathbf{y}_j \right) \right]$$

where components of $\boldsymbol{\lambda}$ are considered as the coordination variables which will be fixed at the upper level, say $\boldsymbol{\lambda} = \overline{\boldsymbol{\lambda}}$, and this allows to decompose the problem (5) into N independent subproblems which can be considered as weighting problems for each of the N subsystems, i.e. the new objectives function $\widetilde{f}^i$ of the subsystem i is obtained by adding the term $T^i = 1/\mu_i \cdot \overline{\boldsymbol{\lambda}}^i)^T \mathbf{x}_i - \sum_{l=1}^N 1/\mu_i(\overline{\boldsymbol{\lambda}}^l)^T \mathbf{C}_{lj}\mathbf{y}_j$ to each component of the objective function.

In [4] a two-level hierarchical algorithm is proposed for the problem (1)-(4) and it is shown that under certain constraints the dual objective is differentiable and the set of properly efficient solutions for the global problem is easily generated by the properly efficient solutions of the subsystems provided they satisfy certain additional so called coordination problem. (Note, that, in general, this assertion is not true.) In this case, the decomposition is carried out by assigning coordination weights to each subsystem μ_i, $i = 1, 2, ..., N$ and fixing the multipliers associated to the coupling $\overline{\boldsymbol{\lambda}} \in^{nx}$, so that N multiobjective subproblems of lower dimension than that of the original problem appear, which will be solved at the lower level. Although the coordinator need not to know an analytic expression of (5) he/she can use the information in order to build a tangential approximation of the weighted function. New values of $\boldsymbol{\lambda}$ can be obtained by maximizing a tangential function and doing so is equivalent to the solution of the system of the nonlinear equations

$$\nabla\Phi(\boldsymbol{\lambda}) = 0, \tag{8}$$

provided that $\Phi(\boldsymbol{\lambda})$ is differentiable. For solving (8) one can use the following iterative procedure

$$\boldsymbol{\lambda}^{s+1} = \boldsymbol{\lambda}^s + \gamma_s d^s, \quad s = 0, 1, ...,$$

where $d^s = \nabla\Phi(\boldsymbol{\lambda}^s)$ and $\gamma_s > 0$ is a properly chosen step-length.

Note, that, the gradient $\nabla\Phi$, in general, does not exist at the solution point $\mathbf{x}(\boldsymbol{\lambda})$ of the problem (5). Then the dual problem could accordingly solved by a non-smooth optimization method

$$\boldsymbol{\lambda}^{s+1} = \max\left\{0, \, \boldsymbol{\lambda}^s + \rho_s \nabla\widehat{\Phi}(\boldsymbol{\lambda}^s)\right\}, \tag{9}$$

where $\nabla\widehat{\Phi}(\boldsymbol{\lambda})$ is subgradient of (5) and ρ_s is chosen such that $\rho_s \to +0$ with $s \to \infty$ and $\sum_{s=0}^{\infty} \rho_s = \infty$ [5],[7].

Although, gradient methods are, in principle, plain to use they suffer from serious disadvantages: in the case of ill-conditioning they require a huge number of iterations to achieve the solution with the prescribed accuracy or they even do not assure the prescribed accuracy at all. In order to reduce the condition number preconditioning techniques turn to be useful. But the situation may be more bad when the problem (8) is ill-posed and one has to seek for a generalized solution in the sense of least squares.

3. Parameter identification

Many kinds of system can be described by optimization models. Usually we do not have a ready-made optimization model waiting for a solution. Applied modelling is based on simplifications and invalid model does not provide a valid answer. A mathematical model in use has to be not only qualitatively correct but it should be numerically sufficiently accurate as well. An important problem in modelling is that of parameter identification from available performance data. Constructing an appropriate mathematical model by experiments and observations involves as a rule the solution of ill-posed problem.

Modelling based on economical, technical and environmental criteria include many sources of uncertainty and subjectivity. This matter has stimulated a study of new, robust and heuristics concepts of approximation and optimization. Genetic algorithms and artificial networks are reliable tools for identification of complex non-linear systems and processes based on input-output representation of the model. A trained feed-forward network is capable of approximating an unknown mapping herewith a proper application dependent architectures of the network can significally improve the quality of identification.

According to the Stone-Weierstrass theorem an arbitrary continuous function of N variables can be approximated by a three-layer-perceptron-like network with $N(2N+1)$ neurons using a continuously increasing activation function. In most applications activation function is sigmoidal one, typically of the form $\frac{1}{1-e^{-x}}$ or hyperbolic tangent function.

The main advantage of using sigmoid functions is that they are everywhere differentiable and it is very easy and fast to calculate the derivatives of these functions. That is why it is desirable to use multi-layer-feed-forward neural networks with Levenberg-Marquardt–type learning rule which is much faster and reliable than the traditional steepest descent. Usually mean square error is used as the criteria for adjusting the weights to minimize the classification error of the network of neurons which leads to solving of highly non-linear least squares problems.

Thus solving non-linear least squares problems is vital for finding model parameters as well as coordination parameters in decomposition-coordination schemes.

4. Numerical methods

Futher on, we shall reformulate the problem (8) into the form more customary for mathematics

$$F(x) = 0, \tag{10}$$

where $F \equiv \nabla\Phi$ and x will stand for argument and discuss some numerical methods for solving nonlinear equations and parameter identification based input-output data.

To make effective decisions, it is important to be aware of existence and changing nature of the potential parameters. Parameters like weights are often difficult to estimate and sometimes require repeated interaction between the methods under consideration. Frequently mean squares error is used as the minimization criterion for adjusting the weights that leads to minimizing a functional of the type

$$f(x) = \frac{1}{2}\left\|F(x)\right\|^2, \tag{11}$$

where $F(x)$ is acting between Hilbert spaces H_1 and H_2.

While the evaluation of proper model parameters and coordination parameters may lead to the solution of complicated systems of nonlinear equations and/or nonlinear least squares problems then sophisticated algorithms are needed for solving them in order to find a trade-off between robustness, stability and efficiency. Thus, the solution of the system of nonlinear equations is a crusial problem in decomposition-coordination schemes.

High order methods for decomposition-coordination problems were discussed in [8], [9]. Here we shall consider methods based on the Gauss-Newton method which can be used for solving either systems of nonlinear equations or least squares problems.

If the operator F is highly non-linear then one possibility to handle non-linear least squares (NLSQ) problems with small residual more safely is to use damped Gauss-Newton–type methods. On the other hand, in the case of ill-conditioned F', acceptable values for the relaxation parameter ε can be extremely small what, in turn, means that the convergence speed may be drastically slowed down. The use of some iterative schemes that exploit the variable regularization parameter α and the variable relaxation parameter ε at each iteration step may be fruitful because that allows a wider choice of initial guesses x_0 and to improve the convergence rate [12].

In case of the bounded pseudoinverse of F' we shall consider methods of the type

$$x_{k+1} = x_k - \varepsilon_k A_k F(x_k), \quad k = 0, 1, \dots, \tag{12}$$

where ε_k is a relaxation (damping) parameter $0 < \varepsilon_k \leqslant 1$ and A_k is an approximation to the IU-pseudoinverse operator $U^{-1}(F'(x)U^{-1})^{+}$ and U is a linear nonsingular operator from H_1 onto H_1 [10], [11].

Putting $H = U^*U$, $B_k = [F'(x_k)]^* F'(x_k)$, $M_k = B_k + \alpha_k H$ with $\alpha_k > 0$ and $A_k = M_k^{-1}[F'(x_k)]^*$ then (12) turns to be a Levenberg-Marquardt (regularized Gauss-Newton) type method

$$x_{k+1} = x_k - \varepsilon_k M_k^{-1}[F'(x_k)]^* F(x_k), \tag{13}$$

where $[\cdot]^*$ denotes the dual mapping.

The family of methods (12) includes also such approximate variants of (13)

$$x_{k+1} = x_k - \varepsilon_k D_k [F'(x_k)]^* F(x_k), \tag{14}$$

where D_k is an approximation of M_k^{-1}, $A_k = D_k[F'(x_k)]^*$ and $\|I - M_k D_k\| \leqslant \mu < 1$. In particular, the operator D_k might be generated by the formula

$$D_k = D_k^{(0)}[I + T_k + \dots + T_k^q], \quad q \geqslant 1, \tag{15}$$

where $T_k = I - M_k D_k^{(0)}$ and in the capasity of $D_k^{(0)}$ one might take, for instance, $D_k^{(0)} = \tau_k I$ with $\tau_k > 0$. Some convergence results for the methods (13) can be found in [10], [11]. Note, that for $q = 0$ and $\varepsilon_k = \tau_k$ the methods (13) - (14) appears to be a steepest descent type method

$$x_{k+1} = x_k - \varepsilon_k g(x_k), \quad \text{with} \quad g(x_k) = [F'(x_k)]^* F(x_k), \tag{16}$$

while in the limiting case $q \to \infty$ it approaches to the Ben-Israel method involving the computation of the pseudoinverse operator. In the case $\alpha_k = 0$ and $q = 2$ the two parametric method

$$x_{k+1} = x_k - \varepsilon_k \left\{ \tau_k [F'(x_k)]^* - \tau_k [F'(x_k)]^* (I - \tau_k F'(x_k)[F'(x_k)]^*) \right\} F(x_k) \tag{17}$$

follows from (13), (15) for which the total number of arithmetic operations per iteration is $O(n^2)$. For sparse Jacobian the amount of computational work can be perceivably diminished which is not insignificant for large-scale problems.

5. Computational aspects

To get more realistic impression of convergence properties of the methods under discussion their approximate variants are here studied. Frequently, the use of finite difference approximations to the derivatives gives rise to an inexact method. An approximate variant of the method can also be obtained as a result of a strategy used for solving linear problems at each iteration step.

Performance of methods of the type (12) is equivalent to either solving the associated linear equations or computing inverses with an error at every iteration step. A strategy of problem solving that instead of finding the exact solution of a linear equation at every step solves it intentionally inexactly permits to save the computational work and is adaptive in the sense that one uses low accuracy numerical solutions of linear equations when the solution of the primary problem is not reached yet and improves the accuracy as the solution is approached. In many cases iterative methods are more appropriate and economical for linear problems than direct ones. Besides, iterative methods are self-correcting, and hence they are not sensitive to computational errors. But their convergence can be quite slow in the presence of ill-conditioning. In order to improve the stability and to save thereby laborious solving of linear auxiliary problems it might be fruitful to do some extra computational effort and implement preconditioning techniques. The basic idea of preconditioning is to introduce a preliminary scaling of the independent and dependent variables. In particular, the problems of polynomial preconditioning can be found in the paper [11]. In recent years, there has been a revival of polynomial preconditioning.

The performance and reliability of the different gradient methods vary considerably. Actually, methods (14) with D_k defined by (15) are all based on the steepest descent direction and therefore for small q they are able to converge from relatively poor initial quess but from beginning $q \geqslant 2$ the implementation of the formula (15) is equivalent to the use of polynomial preconditioning techniques to reduce the spectral condition number that enables to alleviate the disadvantages peculiar for gradient type methods. Recall, that in the limiting case $q \to \infty$ the methods (14), (15) approaches to the Ben-Israel method involving the computation of the pseudoinverse operator [12].

— An essential feature of formulas (14) and (15) is their insusceptibility to small perturbations in F' whereas they mainly involve safe arithmetic operations, i.e., Jacobian matrix F' (or F'^*) times a vector that can be calculated as scalar products. The calculation of a scalar product is a nonexpensive operation and on that account the total number of arithmetic operations needed at each iteration is $O(n^2)$.

— As the key operation for polynomial preconditioning [13], [14] in question is a matrix-vector multiplication then it is an attractive technique on vector and parallel computers.

— One possibility to handle some kind equations with nonsmooth functions is to approximate the locally Lipschitzian function with a smooth one and to use the derivative(s) of the smooth function in the algorithm whenever a derivative is needed [7], [13], [14]. An alternative is then to use a polyalgorithmic approach in a way that methods developed for nonsmooth problems will be implemented in regions of nonsmoothness.

6. Concluding remarks

Although we have discussed here the methods only on the theoretical basis, numerical experience with the methods under consideration has also confirmed these theoretical considerations. The numerical results of computational experiments can be partially found in [8], [14]. These promising results encourage us to carry on the investigation of the convergence properties of the methods under discussion and of polyalgorithmic procedures and also exploit more bigger amount of test problems to estimate their actual (practical) convergence rate.

Multi-objective optimization results in a number of solutions and there is no single best solution but a set of solutions nevertheless the final stage of multi-objective optimization consists of the selection of single solution. In selecting the best instrument to decision making it is necessary to find a solution that gives the best outcome with respect to objectives of all units (in terms of sustainability for all society members).

СПИСОК ЛИТЕРАТУРЫ

1. *Miettinen, K.* On the methodology of multiobjective optimization with applications. University of Jyväskylä. 1994

2. *Ulm, S.* Decomposition methods for the solution of optimization problems. Tallinn. 1979 (in Russian)

3. *Lootsma, F.A.* Exploitation of structure in non-linear optimization. In Parallel Computing 89, D. J. Evans et al., eds., North–Holland. 1990. P. 31–45

4. *Miguel, F. et al.* A decomposition–coordination method for complex multi-objective systems. Asia-Pasific Journal of Optional Research. 26. 2009. P. 735–757

5. *Ermolyev, Y. M. et al.* Mathematical methods for operation research. Kiev. 1979 (in Russian)

6. *Bazaraa, M., Sherali, H. and Shetty, C. M.* Nonlinear programming. Theory and algorithms. 2nd edition, John Wiley & Sons, New York. 1993

7. *Shor, N. Z.* Minimization methods for non-differentiable functions. Springer-Verlag, Berlin. 1985

8. *Vaarmann, O.* On high order iterative methods for nonlinear problems. In: Nonlinear Mathematical Analysis and Applications, (Th.M.Rassias ed.). Palm Harbour, Fl, Hadronic Press. 1998. P. 275–299

9. *Vaarmann, O.* On high order iterative methods for decomposition–coordination problems. Technological and Economic Development of Economy. 2006. 12, P. 56–61

10. *Vaarmann, O.* Solution of nonlinear least squares problems by Levenberg-Marquardt type methods. Proceedings of the Estonian Academy of Sciences: Physics, Mathematics. 1989. 38, No 2, P. 146–153

11. *Vaarmann, O.* On solving ill-conditioned systems of nonlinear equations. Proceedings of the Estonian Academy of Sciences: Physics, Mathematics. 1994. 43, No 2, P. 49–63

12. *Kangro, I. and Vaarmann, O.* Some iterative regularized methods for highly nonlinear least squares problems. Mathematical Modelling and Analysis. 2009. 14, No 2, P. 179–186

13. *Miller, G. L. and Richter, P. C.* Lower bounds for graph embeddings and combinatorial preconditioners. in SPAA '04: Proceedings of the sixteenth annual ACM symposium on Parallelism in Algorithms and Architectures, ACM. 2004. P. 112–119

14. *Fisher, B. and Freund, R.* On adaptive weighted polynomial preconditioning for hermitian positive definite matrices. SIAM J. Sci. Comp. 1994. 15, P. 408–426

15. *Qi Li Qun* Trust region algorithms for solving nonsmooth equations. SIAM J.Optim. 1995. 5, P. 408 – 426

16. *Gabriel, A. and Jong-Shi Pang* A trust region methods for constrained nonsmooth equations. Large Scale Optimization: State of the Art (eds. W.W. Hagerer et al), Kluwer Academic Publishers. Dordrecht. 1994. P. 155–181

17. *Amat, S., Busquier S., Gutierrez, J. M.* On the local convergence of secant type methods. Intern. J. Comp. Math. 2004. 81, P. 1153–1161